McDermott
Will & Emery

HIPAA De-Identification Guidance

December 11, 2012
Office of Civil Rights has released additional guidance addressing the de-identification of protected health information in accordance with the HIPAA Privacy Rule. Covered entities should review their current de-identification methods and make any necessary changes to comply with the new guidance.

On November 26, 2012, the Office for Civil Rights (OCR) released guidance regarding methods for de-identification of protected health information (PHI) in accordance with the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule (Privacy Rule).

The guidance largely restates prior interpretive guidance to and health care industry understandings of the Privacy Rule's de-identification standard. Since the guidance follows a lengthy process of public meetings and other opportunities for input from stakeholders, it appears that OCR has determined that the current de-identification standard strikes an appropriate balance between individuals' interest in the privacy of their personal information and the interests of the research community and other data users. For more information about OCR's proposed modifications to the Privacy Rule, see McDermott's *White Paper* "OCR Issues Proposed Modifications to HIPAA Privacy and Security Rules to Implement HITECH Act."

Background

The Privacy Rule applies to PHI, which is individually identifiable health information (subject to certain limited exceptions). Individually identifiable health information is defined as follows:

- Information created or received by a health care provider, health plan, employer or health care clearinghouse
- Information that relates to the past, present or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present or future payment for the provision of health care to an individual
- Information that identifies the individual, or with respect to which there is a reasonable basis on the part of the disclosing entity for believing that the information may be used to identify the individual

The HIPAA Privacy Rule provides a pathway for covered entities and other health data users to create and then use and disclose de-identified health information outside the disclosure restrictions on PHI. De-identified information is health information that does not identify an individual, and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual.

The Privacy Rule establishes two methods for a covered entity to de-identify information: (1) obtaining a professional statistical analysis and opinion regarding de-identification; or (2) removing 18 specific identifiers.

Removal of 18 Specific Identifiers Method

Information is deemed to be de-identified if all of the following identifiers of the individual or of relatives, employers or household members of the individual are removed, and the covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information:

- Names
- All geographic subdivisions smaller than a state, including street address, city, county, precinct, ZIP code and their equivalent geocodes, except for the initial three digits of a ZIP code if, according to the current publicly available data from the Bureau of the Census, (1) the geographic unit formed by combining all ZIP codes with the same three initial digits contains more than 20,000 people, and (2) the initial three digits of a ZIP code for all such geographic units containing 20,000 or fewer people are changed to 000

State Administration & Veterans' Affairs
November 17, 2015
Exhibit 20

- All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date and date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of "age 90 or older"
- Telephone numbers
- Fax numbers
- E-mail addresses
- Social Security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Device identifiers and serial numbers
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) addresses
- Biometric identifiers, including finger and voice prints
- Full-face photographic images and any comparable images
- Any other unique identifying number, characteristic or code

Professional Statistical Analysis

Information will be deemed to be de-identified for HIPAA compliance purposes if a person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable takes the following actions:

- Applies such principles and methods, and determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information
- Documents the methods and results of the analysis that justify such determination

Covered entities, business associates and other data users often choose the Professional Statistical Analysis approach (and incur the professional's fees) instead of relying upon the Removal of 18 Specific Identifiers approach, because the professional may issue an opinion that allows certain of the 18 identifiers to be included in the de-identified data set.

General Guidance

The guidance reaffirms the long-held understanding that a covered entity may engage a business associate to de-identify PHI on the covered entity's behalf—for example, if the covered entity does not have the experience or resources to perform the data scrubbing. The guidance stresses, however, that the business associate agreement must expressly authorize the business associate to perform this activity. Thus, in light of this guidance, business associate agreements that refer generally to health care operations may not be sufficient to direct the business associate to perform de-identification services.

Additional Guidance with Respect to the Removal of 18 Specific Identifiers Method

The guidance provides additional details with respect to the Removal of 18 Specific Identifiers Method. Below are summarized some of the relevant provisions.

May parts or derivatives of any of the listed identifiers be disclosed consistent with the Removal of 18 Specific Identifiers Method?

- No. For example, a data set that contained patient initials or the last four digits of a Social Security number would not meet the requirement of the Removal of 18 Specific Identifiers Method for de-identification.

What are examples of dates that are not permitted according to the Removal of 18 Specific Identifiers Method?

- Elements of dates that are not permitted for disclosure include the day, month and any other information that is more specific than the year of an event. For instance, the date January 1, 2009, could not be reported at this level of detail. However, it could be reported in a de-identified data set as 2009.
- Many records contain dates of service or other events that imply age. Ages that are explicitly stated or implied as over 89 years old must be recoded as 90 or above. For example, if the patient's year of birth is 1910 and the year of health care service is reported as 2010, then in the de-identified data set the year of birth should be reported as "on or before 1920." Otherwise, a recipient of the data set would learn that the age of the patient is approximately 100.

Can dates associated with test measures for a patient be reported in accordance with Safe Harbor?

- No, except as provided above.

What constitutes "any other unique identifying number, characteristic or code" with respect to the Removal of 18 Specific Identifiers Method of the Privacy Rule?

- This category corresponds to any unique features that are not explicitly enumerated in the Safe Harbor list (A–Q) but could be used to identify a particular individual. Examples include indentifying numbers, codes or characteristics.

What is "actual knowledge" that the remaining information could be used either alone or in combination with other information to identify an individual who is a subject of the information?

- The guidance provides that "actual knowledge" means clear and direct knowledge that the remaining information could be used, either alone or in combination with other information, to identify an individual who is a subject of the information. This means that a covered entity has actual knowledge if it concludes that the remaining information could be used to identify the individual. The covered entity, in other words, is aware that the information is not actually de-identified information.

Must a covered entity suppress all personal names, such as physician names, from health information for it to be designated as de-identified?

- No. Only names of the individuals associated with the corresponding health information (*i.e.*, the subjects of the records) and of their relatives, employers and household members must be suppressed.

Must a covered entity remove protected health information from free text fields to satisfy the Removal of 18 Specific Identifiers Method?

- The guidance notes the risk associated with contextual identifiers in free text and other unstructured data fields (such as physician progress notes of a medical record). When relying on the removal of the 18 identifiers to achieve de-identification, covered entities should take special care to ensure that unstructured data fields do not contain stray identifiers (for example, a hand-written name on an x-ray scan) or information that could be used to re-identify the patient (such as noteworthy professional or athletic roles or accomplishments).

Must a covered entity use a data-use agreement when sharing de-identified data to satisfy the Removal of 18 Specific Identifiers Method?

- No. As stated above, the Privacy Rule does not limit how a covered entity may disclose de-identified health information. However, the guidance notes that a covered entity may require the recipient of de-identified information to enter into a data-use agreement. A covered entity should enter into such a use agreement to address intellectual property ownership issues (such as who owns the de-identified data set) and any business concerns regarding the purposes for which the data set may be utilized.

It is also noteworthy that the guidance does not address the emerging question of whether genetic information is an example of a "unique code" under the 18th identifier.

Additional Guidance with Respect to the Professional Statistical Analysis Approach

The guidance provides additional details with respect to the Professional Statistical Analysis approach. Most of this guidance is directed towards the "expert" chosen by the covered entity. Below are summarized some of the relevant provisions.

Who is an "expert?"

- The guidance provides that there is no specific professional degree or certification program for designating who is an expert at rendering health information de-identified. Suggested experts include individuals with statistical, mathematical or other scientific backgrounds. From an enforcement perspective, OCR would review the relevant professional experience and academic or other training of the expert used by the covered entity, as well as actual experience of the expert using health information de-identification methodologies.

What is an acceptable level of identification risk for an expert determination?

- The guidance states that there is no explicit numerical level of identification risk that is deemed to universally meet the "very small" level indicated by the method. The analysis is more of a facts and circumstances analysis based on the ability of a recipient of information to identify an individual (i.e., subject of the information). This is notable as it preserves a degree of latitude for statistical experts engaged to de-identify information to place "very small risk" into context informed by any number of relevant factors, including the specific intended recipient. It also demonstrates that OCR recognizes that a "very small" risk of re-identification is not the same as no risk, and that covered entities are not out of compliance if re-identification occurs despite the statistical expert's expectation that it would not.

How long is an expert determination valid for a given data set?

- There is no *per se* expiration date. The guidance does, however, state that experts recognize that technology, social conditions and the availability of information changes over time. For example, the U.S. Department of Commerce's release of U.S. census data may affect the ongoing validity of a statistical opinion. Thus, experts should assess the expected change of computational capability, as well as access to various data sources, and then determine an appropriate timeframe within which the health information will be considered reasonably protected from identification of an individual. Covered entities and others requesting statistical opinions should expect the expert to request that the statistical opinion only be valid for a certain length of time and factor in the cost of renewals of the opinion when deciding whether to pursue the Professional Statistical Analysis over the Removal of 18 Specific Identifiers Method.
- Information that had previously been de-identified may still be adequately de-identified when the certification limit has been reached. When the certification timeframe reaches its conclusion, it does not imply that the data that has already been disseminated is no longer sufficiently protected in accordance with the de-identification standard. Covered entities will be obliged to have an expert examine whether future releases of the data to the same recipient (*e.g.*, monthly reporting) should be subject to additional or different de-identification processes consistent with current conditions to reach the very low risk requirement.

How do experts assess the risk of identification of information?

- The guidance provides that there is no single universal solution that addresses all privacy and identifiability issues. The guidance suggests that a combination of technical and policy procedures be applied to the de-identification task. A sample workflow for expert determination is depicted in the guidance in the form of a flowchart. In addition, a sample chart is provided to demonstrate the principles used by experts in the determination of the identifiability of health information.
- The guidance recognizes that the Professional Statistical Analysis is an iterative process that takes into account a number of factors. For example, one might expect that specific details regarding the covered entity, the covered entity's data co-mingling systems, the data recipient, the data itself and many other factors would inform the judgment. This underscores that it is not just the specific data fields that are included that inform whether information is de-identified, but also the entire data-sharing arrangement. It also suggests that a covered entity might require multiple statistical opinions to govern different data-sharing arrangements and that a data set deemed de-identified in one context might remain identifiable in another, even within the same covered entity. Covered entities should consider whether the expert should document the range of circumstances under which the opinion is valid.

What are the approaches by which an expert assesses the risk that health information can be identified?

There is no bright line rule. The de-identification standard does not mandate a particular method for assessing risk, but it does provide a survey of potential approaches.

Must a covered entity use a data-use agreement when sharing de-identified data to satisfy the Expert Determination Method?

No. The Privacy Rule does not require a covered entity to enter into a data-use agreement in order to share a de-identified data set. However, as noted above, it is recommended that a covered entity should enter into a data-use agreement to address intellectual property ownership issues (such as who owns the de-identified data set) and business concerns regarding the purposes for which the data set may be utilized.

Next Steps

Covered entities (and business associates with the right to de-identify PHI that they receive from their customers) should review their current de-identification methods in light of the guidance and make any necessary changes to comply with the new guidance. As part of the review, data users should consider whether a previously issued opinion needs to be refreshed in light of new publicly available data sources, such as census data. If you have any questions, contact your regular McDermott Will & Emery lawyer or one of the contacts listed to the right for assistance.

Authors

Jennifer S. Geetter

Amy M. Gordon

Daniel F. Gottlieb

Amy Hooper Kearbey

Practice Areas & Industries:

Employee Benefits

Health

HIPAA Privacy and Security Solutions

Subscribe

Subscribe to Newsletters and News Alerts

Follow @McDermottLaw

Follow Us